https://helda.helsinki.fi

An assessment of ten ocean reanalyses in the polar regions

Uotila, Petteri

2019-02

Uotila , P , Goosse , H , Haines , K , Chevallier , M , Barthelemy , A , Bricaud , C , Carton , J , Fuckar , N , Garric , G , Iovino , D , Kauker , F , Korhonen , M , Lien , V S , Marnela , M , Massonnet , F , Mignac , D , Peterson , K A , Sadikni , R , Shi , L , Tietsche , S , Toyoda , T , Xie , J & Zhang , Z 2019 , ' An assessment of ten ocean reanalyses in the polar regions ' , þÿ C l i m a t e D y n a m i c s , v o l . 5 2 , n o . 3 - 4 , p p . 1 6 1 3 1 6 5 0 . h t t p s : / / d o i . o r g /

http://hdl.handle.net/10138/300090
https://doi.org/10.1007/s00382-018-4242-z

cc_by
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Climate Dynamics
https://doi.org/10.1007/s00382-018-4242-z

An assessment of ten ocean reanalyses in the polar regions

Petteri Uotila[1] • Hugues Goosse[2] • Keith Haines[3] • Matthieu Chevallier[4] • Antoine BarthØlemy[2] • ClØment Bricaud[5] • Jim Carton[6] • Neven Fu kar[7,8] • Gilles Garric[5] • Doroteaciro Iovino[9] • Frank Kauker[10] • Meri Korhonen[11] • Vidar S. Lien[12] • Marika Marnela[11] • François Massonnet[2,7] • Davi Mignac[3] • K. Andrew Peterson[13] • Remon Sadikni[14] • Li Shi[15] • Ste en Tietsche[16] • Takahiro Toyoda[17] • Jiping Xie[18] • Zhaoru Zhang[19]

Received: 11 October 2017 / Accepted: 17 April 2018
' The Author(s) 2018

Abstract

Global and regional ocean and sea ice reanalysis products (ORAs) are increasingly used in polar research, but their quality remains to be systematically assessed. To address this, the Polar ORA Intercomparison Project (Polar ORA-IP) has been established following on from the ORA-IP project. Several aspects of ten selected ORAs in the Arctic and Antarctic were addressed by concentrating on comparing their mean states in terms of snow, sea ice, ocean transports and hydrography. Most polar diagnostics were carried out for the rst time in such an extensive set of ORAs. For the multi-ORA mean state, we found that deviations from observations were typically smaller than individual ORA anomalies, often attributed to o - setting biases of individual ORAs. The ORA ensemble mean therefore appears to be a useful product and while knowing its main de ciencies and recognising its restrictions, it can be used to gain useful information on the physical state of the polar marine environment.

Keywords Oceanography • Reanalyses • Arctic • Antarctic • Sea-ice

Marika Marnela was formerly at Finnish Meteorological Institute.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00382-018-4242-z) contains supplementary material, which is available to authorized users.

Petteri Uotila
petteri.uotila@helsinki.

1 Institute for Atmospheric and Earth System Research (INAR)/Physics, University of Helsinki, Helsinki, Finland

2 Earth and Life Institute, Universitø Catholique de Louvain, Louvain-la-Neuve, Belgium

3 University of Reading and National Centre for Earth Observation, Reading, UK

4 Centre National de Recherches MØtØorologiques, MØtØo France/CNRS UMR3589, Toulouse, France

5 Mercator OcØan, Toulouse, France

6 University of Maryland, College Park, USA

7 Barcelona Supercomputing Centre, Barcelona, Spain

8 Environmental Change Institute, University of Oxford, Oxford, UK

9 Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

10 Alfred Wegener Institute, Bremerhaven, Germany

11 Finnish Meteorological Institute, Helsinki, Finland

12 Institute of Marine Research, Bergen, Norway

13 Met O ce, Exeter, UK

14 University of Hamburg, Hamburg, Germany

15 Bureau of Meteorology, Melbourne, Australia

16 European Centre for Medium-Range Weather Forecasts, Reading, UK

17 Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

18 Nansen Environmental and Remote Sensing Center, Bergen, Norway

19 Shanghai Jiao Tong University, Shanghai, China

1 Introduction

For years, atmospheric reanalysis products, which consist of multidecadal meteorological model simulations with assimilated observations, have become an invaluable resource for researchers representing a wide range of disciplines. Recently, similar products ocean reanalyses (ORAs) have been constructed by many research groups. It is likely that these products will become as valuable as their atmospheric counterparts.

Speci cally, an ocean analysis describes an ocean state valid for a particular time by a set of gridded oceanographic variables. Typically an ocean analysis is generated by an analysis system consisting of a hydrodynamical or statistical model and an observation assimilation framework, for the purpose of initialising a forecast. During the analysis generation process, the forecast model background state is adjusted toward new observations. The amount of adjustment is denoted as the analysis increment, which quantify the impact of data assimilation in the analysis system (Cullather and Bosilovich 2012).

Ocean and sea ice reanalyses are analyses in the form of time series, where every analysis is generated using the same analysis system for all historical observations. Hence, they combine observations either statistically or with a hydrodynamical model, to reconstruct historical conditions and their changes in the ocean.

Global and regional ORA products are increasingly used in polar research, but their quality remains to be systematically assessed. To address this, the Polar ORA Intercomparison Project (Polar ORA-IP) has been established following on from the ORA-IP project (Balmaseda et al. 2015; Toyoda et al. 2017a, b; Chevallier et al. 2017; Tietsche et al. 2015; Karspeck et al. 2015; Shi et al. 2017; Valdivieso et al. 2017; Palmer et al. 2017; Masina et al. 2015; Storto et al. 2017).

These ORA-IP studies have looked at various aspects of global ocean hydrodynamics (steric sea level, air-sea uxes, ocean heat and salt content among others). The only ORA-IP publication with a polar focus has been Chevallier et al. (2017), who compared the representation of the sea-ice cover in the Arctic Ocean in 14 global reanalyses. Using a variety of in-situ and satellite-based observational datasets, they investigated mean states, trends and interannual variability in these reanalyses, focusing on sea-ice concentration (with extent and area), thickness (with volume), velocity and snow depth over sea ice.

Chevallier et al. (2017) showed consistency with respect to sea-ice concentration, which is primarily due to the constraints in surface temperature imposed by atmospheric forcing, and ocean-ice data assimilation. However, they found a large spread in sea-ice and snow thicknesses within the ensemble of ORAs, due to biases in the ocean-ice model components, and lack of observational constraint. Chevallier et al. (2017) discussed the possible role of model parameters, prescribed atmospheric forcing and data assimilation on the spread. They concluded that none of the ORAs stands superior to the others when compared with observed sea-ice thickness calculated from satellite altimetry data, and that data assimilation does not seem to improve the simulated sea-ice thickness. As a result, estimates of Arctic sea-ice volume by individual ORAs su er large uncertainties, and the ORA multi-model ensemble mean (MMM) ice volume does not provide a more robust estimate. Most of the global reanalyses used in Chevallier et al. (2017) have now been updated and their updates are evaluated in the present paper which allows direct comparisons with their results.

In this study, we aim for a comprehensive evaluation of ten selected ORA products (C-GLORS025v5, ECDA3, GECCO2, Glorys2v4, GloSea5-GO5, MOVE-G2i, ORAP5, SODA3.3.1, TOPAZ4 and UR025.4) in the Arctic and Southern Oceans (Table 1). For these regions the diagnostics target the following topics: hydrography; ocean heat (OHC), salt content (OSC); ocean transports; mixed layer depth (MLD); sea-ice concentration (SIC) and thickness (SIT); and snow thickness over sea ice. The ORA product biases against observed reference data and their mutual spread are quanti ed, and possible reasons for discrepancies discussed.

The scope of our manuscript is to provide a broad state-of-the-science overview of ocean reanalyses, plus our best estimate of what the truth might look like. In this context, we will check if the MMM is a useful estimate. As we will repeatedly show, it is a set of elds which is generally most consistent with observations. This is what many users require, although it may not be best suited to analysing dynamical or physical processes, for example.

If a user does not want the MMM, but would prefer a single ORA output, for instance to understand the dynamics, this paper does not seek to tell the user which one to use, but in addition to providing a general evaluation, it is able to show which are outliers for certain variables, which can still be very useful.

We pay particular attention to the performance of the MMM compared to individual products and the identi cation of outliers. Notably, as the ORAs assimilate observations they are not independent of some of the reference data they are compared to. Moreover, we investigate links and covariability between the diagnostics, such as the Arctic Ocean heat content and North Atlantic heat transport, and between the mixed layer depth, oceanic convection, the upper ocean hydrography, sea ice and snow. In this way, we try to identify physical mechanisms causing common and individual ORA biases.

Although a large majority of the existing ORA publications does not focus on polar regions, the Coordinated Ocean Reference Experiment (CORE-II; Danabasoglu et al.

Table 1 List of ten ocean reanalyses used in the study and their central characteristics

Name	C-GLORS025v5	ECDA3	GECCO2	GLORYS2v4	GloSea5-GO5	MOVE-G2i	ORAP5	SODA3.3.1	TOPAZ4	UR025.4
Institution	CMCC	GFDL/NOAA	Hamburg University	Mercator Oc∅an	UK MetO ce	MRI/JMA	ECMWF	University of Maryland	NERSC	University of Reading
Nominal horizontal resolution									12 16 km	
Vertical resolution	50 z-levels	50 z-levels	50 z-levels	75 z-levels	75 z-levels	52 z-levels	75 z-levels	50 z-levels	28 z-isopycnal layers	75 z-levels
Top-level thickness	1 m	10 m	10 m	1 m	1 m	2.25 m	1 m	10 m	min 3 m	1 m
Ocean-ice model	NEMO3.2-LIM2	MOM4-SIS	MITgcm	NEMO3.1-LIM2	NEMO3.4-CICE	MRI.COM3-CICE4	NEMO3.4-LIM2	MOM5-SIS	HYCOM-EVP SI	NEMO3.2-LIM2
Time period	1980 2015	1961 2012	1948 2014	1992 2015	1993 2012	1980 2012	1979 2012	1980 2015	1991 2016	1989 2010
Initialization	Spinup	Spinup	Cold start	Cold start	Spinup	Spinup	Spinup	Spinup	Cold start	Cold start
Source of atmospheric forcing data	ERA-Interim	Coupled	NCEP RA1	ERA-Interim	ERA-Interim	JRA-55[a]	ERA-Interim	NASA MERRA2	ERA-Interim	ERA-Interim
Ocean restoring	Large scale bias correction to EN3v2a	Fully coupled	None	T, S restoring towards EN4.1.1 for z > 2000 m and lat < 60 S (= 20 years)	Surface Haney SSS restoring (33.333 mm/day/PSU), 3D T/S to ENACT3 2004 2008 climatology (= 1 year)	Relaxing (by IAU) T/S to merged PHC3-WOA13 climatology (= 5 years)	Relaxation to OSTIA/NOAA OIv2d SST	Restoring to mean T and S (= 10 years). Relaxation to WOA SSS (= 3 months)	Relaxing T/S to merged PHC3 WOA13 climatology	None
Sea-ice DA method	Nudging	None (SST)	None (SST)	Reduced order KF	3DVAR	3DVAR	3DVAR-FGAT	None (SST)	EnKF	OI
Sea-ice DA variables	SIC, Arctic SIT			SIC	SIC	SIC	SIC		SIC, SIV	SIC
Sea-ice DA sources	NOAA OIv2d, PIOMAS			CERSAT	OSISAFv2	MGDSST	OSTIA, NOAA OIv2d		OSISAF	OSISAF
Ocean DA method	3DVAR	EnKF	4DVAR (adjoint)	Reduced order KF + 3DVAR large scale bias correction to in-situ T, S	3DVAR	3DVAR	3DVAR	OI	EnKF	OI
Ocean DA variables	T, S, SSH, SST	T, S, SST	T, S, SSH, SST	T, S, SSH, SST	T, S, SST, SSH	T, S, SSH, SST	T, S, SSH, SST	T, S, SST	T, S, SSH, SST	T, S, SSH, SST, SSS

Table 1 (continued)

Name	C-GLORS025v5	ECDA3	GECCO2	GLORYS2v4	GloSea5-GO5	MOVE-G2i	ORAP5	SODA3.3.1	TOPAZ4	UR025.4
Ocean DA sources	EN3v2a, AVISO	HadSST, OISST, WOD09, GTSPP, Argo	EN3v2a, AVISO, GOCO, HadISST, AMSRE, WOA09	CMEMS, NOAA AVHRR	EN4, ICOADS, AVHRR, ATSR, AMSRE, AVISOv3	WOD13, GTSPP, AVISO, MGDSST	EN3v2a, AVISO, OSTIA	WOD, ICOADS, AVHRR, Metosat SEVIRI	NOAA Reynolds, OSTIA, CLS, Damocles	EN3v2a, OSTIA, AVISO
Reference	Storto et al. (2016)	Chang et al. (2013)	Köhl (2015)	Garric et al. (2018)	Blockley et al. (2014, 2015)	Toyoda et al. (2016)	Zuo et al. (2015), Tietsche et al. (2015)	Carton and Giese (2008)	Xie et al. (2017)	Valdivieso et al. (2014)

T temperature data, S salinity, SST sea surface temperature, SSS sea surface salinity, SSH sea surface height, SIC sea-ice concentration, SIT sea-ice thickness, SIV sea-ice velocity, DA data assimilation, KF Kalman Iter, EnKF ensemble Kalman Iter, OI objective interpolation

[a]Climatological radiation biases from GEWEX3.0 (Stackhouse Jr et al. 2011) are corrected

2014) has produced papers (Downes et al. 2015; Farneti et al. 2015; Wang et al. 2016a, b) which evaluate the polar performance of a number of state-of-the-science global ocean models. The main di erence between the CORE-II model con gurations and the ORAs is that the latter employ advanced data assimilation schemes using mostly the same ocean-ice observations, while CORE-II models only apply simple surface ux corrections that, for example, nudge their sea surface salinities toward climatological values. However the CORE-II protocol requires the participating modelling groups to use common atmospheric states and boundary layer parameterisations to drive their multidecadal simulations (e.g. Gri es et al. 2009; Danabasoglu et al. 2014), which is not the case for the ORAs. It is interesting to compare the relative e ectiveness of the common CORE-II framework with the ORA observations in producing consistent results.

Due to these dependencies, comparisons between CORE-II and ORA results potentially enable us to estimate the role of di erent factors a ecting the multi-model skill in the polar oceans. Similarities between CORE-II and the ORA MMM performance may reveal common issues in model physics and resolution, while discrepancies may provide information on the role of data assimilation and atmospheric forcing.

Along with CORE-II results, other relevant literature for the Arctic and Southern Oceans are discussed in the next two Sects. 2.1 and 2.2, respectively. In Sect. 3, we describe our diagnostic methods and in Sect. 4 we represent the analysis results of ten ORAs. These results are then compared with previous results, including Chevallier et al. (2017) and CORE-II, in the discussion (Sect. 5). Conclusions follow in Sect. 6.

2 Observed and simulated changes in the polar oceans

2.1 The Arctic Ocean

The Arctic sea ice has shown an unprecedented decline since the mid-1990s, which also has impacted the state of the Arctic Ocean (Comiso 2012; Polyakov et al. 2013; IPCC 2013; Polyakov et al. 2017). This dramatic change highlights the need for more comprehensive environmental data to assess the state and impacts of the Arctic in transition. However, even after a number of targeted eld expeditions and improved satellite coverage, the Arctic Ocean observations remain sparse compared to the northern North Atlantic. An important reason for this is that with a few exceptions there are no Argo-buoy deployments north of 70 N to provide hydrographic observations, as the buoys cannot operate under perennial sea ice. Furthermore, international research

teams have had restricted access to the observations from the Russian Arctic which has further limited the observational coverage. Climate models appear too conservative in terms of simulating the observed Arctic sea-ice decline, although there have been some improvements, while their prediction accuracy is signi cantly limited by the relatively large climate variability (Stroeve et al. 2012; Jahn et al. 2016; Melia et al. 2015).

Despite the aforementioned limitations, signi cant progress in understanding of the physical state and evolution of the Arctic Ocean has been gained during the last decade. We brie y list some research e orts closely related to the development of ocean reanalysis products in the Arctic.

The Arctic Ocean Model Intercomparison Project (AOMIP) and its successor, the Forum for Arctic Modeling and Observational Synthesis (FAMOS), have in the last two decades identi ed many model shortcomings and come up with recommendations to reduce the impacts of these shortcomings (Proshutinsky et al. 2016). AOMIP and FAMOS have covered a wide range of topics from Arctic Ocean energetics to sea-ice dynamics (for example Uotila et al. 2006; Heimbach et al. 2010; Karcher et al. 2012). The rst AOMIP phase proved that the co-ordinated community approach is the most e ective way to address the degree of uncertainty of model results. During AOMIP, ocean-ice models with data assimilation were rst introduced to the community (see for example Kauker et al. 2009). Later, FAMOS has been a very productive collaborative e ort by producing more than 60 publications including a special issue in the Journal of Geophysical Research (Proshutinsky et al. 2016). The AOMIP/FAMOS modelling studies document, in addition to their scienti c results, important ORA developments in the polar regions from the reanalysis methodological perspective. However, a systematic diagnostic analysis of ORA products in the Arctic is missing from the AOMIP/FAMOS studies. This is likely due to the relatively late appearance of ORAs, which have a global scope, in contrast to the regional AOMIP/FAMOS one, and to the strong process focus of AOMIP/FAMOS.

In addition to the sea-ice changes mentioned above, the upper Arctic Ocean is freshening and Rabe et al. (2014) were able to identify a freshwater ux trend of 600 – 300 km year from 1992 to 2012. The variability of the Arctic freshwater content correlates well with the atmospheric forcing and can be closely reproduced by the regional coupled sea ice-ocean model North Atlantic Arctic Sea Ice Ocean Model (NAOSIM) simulations (Karcher et al. 2003). Rabe et al. (2014) suggest a high freshwater export through the Fram Strait until the mid-1990s, followed by lower export rates with no trend thereafter, although models may show large di erences in terms of interannual variability of the liquid freshwater through the Fram Strait (Jahn et al. 2012).

Some more recent studies present results from individual polar ocean reanalyses and are worth mentioning here. For example, Xie et al. (2017) analysed multi-decadal ensemble simulations from the regional TOPAZ4 ocean-ice data assimilation system in the Arctic and found that TOPAZ4 performed better with respect to near-surface ocean variables compared to subsurface ocean and sea-ice thickness due to sparse observations. Furthermore, the TOPAZ4 skill improved as the polar observation network became denser. Speci cally, TOPAZ4 has a too cold and di use Atlantic water (AW) layer in the Arctic leading to a cold bias of 0.3 C at around 400 m, while the Barents Sea is too warm and saline. Although, the decadal reduction of TOPAZ4 sea-ice extent is close to the observed, its regional distribution has a dipole bias sea-ice concentration is too low close to the ice edge and too high in the central pack, due to the missing sea-ice heat capacity of TOPAZ4 sea-ice model. Xie et al. (2017) also found that the TOPAZ4 sea ice is too thin, on average.

Lien et al. (2016) applied objective statistical methods to assess the added value of data assimilation in three ocean models, including TOPAZ4, for hydrography, volume and heat transports in the Nordic Seas (the Greenland, Iceland, Norwegian and Seas) and the Barents Sea. They found that both data assimilation and higher model resolution improved the model realism. Speci cally, high model resolution in ocean and atmospheric forcing improved the representation of variables closely related to forcing, such as sea-ice concentration and sea surface temperature. Hydrographic data assimilation had a tendency to reduce hydrographic biases, but its e ect on the liquid ocean transport remained limited (Zuo et al. 2011). Lien et al. (2016) found that the modelled heat transports through the Fram Strait to the Arctic Ocean were within the observational range related to generally realistic looking hydrography and currents.

Recently, a set of multidecadal ocean-ice model hindcasts generated following the CORE-II protocol has provided a wealth of information on the performance of state-of-the-science global ocean-ice models in the Arctic Ocean (Danabasoglu et al. 2014). The CORE-II atmospheric state, including the global warming trend, was used to drive the models for 60 years from 1948 to 2007. In total, CORE-II models were run for 300 years, corresponding to 5 consecutive loops of the 60-year forcing period. Wang et al. (2016a) analysed the sea-ice extent, sources of solid freshwater and the solid freshwater content of CORE-II models in the Arctic focussing on the fth forcing cycle. They found that the models reproduced observed sea-ice variability more consistently than the mean state. The CORE-II MMM sea-ice extent was somewhat smaller than observed, in particular in summer, which resulted in a stronger than observed seasonal cycle. The CORE-II MMM overestimated the winter-to-summer sea-ice retreat rate, related to the negative summer sea-ice extent bias. Models that overestimated the sea-ice thickness,

underestimated the multidecadal decline of the Arctic summer sea-ice cover. On average, the models underestimated the observed sea-ice thinning by a factor of two Wang et al. (2016a) stated.

In terms of hydrography, Ilicak et al. (2016) found that while the CORE-II MMM appears to be relatively close to observations, there is a large inter-model temperature spread in the Arctic Ocean. Speci cally, at intermediate depths, including the warm AW layer, modelled-to-observed temperature di erences were large. The CORE-II MMM had a too cold AW at 400 m whose signal disappeared quickly northward away from the Fram Strait, and an overall cold and fresh bias in the Arctic interior, although its mean freshwater transports through the Arctic gateways appear realistic (Wang et al. 2016b). With respect to individual models, those with too cold intermediate depths have an excessive cold water transport to the Arctic Ocean through the St. Anna Trough, while those models with a warm Arctic have a strong in ow of warm water in the Fram Strait. As with sea ice, the CORE-II models agree on the ocean decadal variability, which is dictated by the common atmospheric forcing, more than they do on the ocean mean state. Following these ndings, Ilicak et al. (2016) point out that the CORE-II ocean-ice models have a too coarse horizontal resolution, typically 1 in latitude, to realistically represent the AW in ow, and the deep water formation and currents originating from the shallow continental shelf regions.

2.2 The Southern Ocean

Over recent decades, the Antarctic sea-ice extent has remained relatively stable but with large interannual variability and a small increasing trend that strongly contrasts with the large decline in the Arctic over the same period (Parkinson and Cavalieri 2012; Maksym et al. 2012). Over the Southern Ocean the westerlies have strengthened and shifted southward, spreading the sea ice northward more e ectively (Marshall 2003; Zhang 2014). Below the surface layer, the temperature has risen while a freshening is observed in many areas (Gille 2008; Schmidtko et al. 2014; de Lavergne et al. 2014). Simulations performed with coupled climate models are generally not able to adequately reproduce these trends. In particular, the majority of them display a decrease in ice extent over the last 30 years in response to anthropogenic forcing. Part of the discrepancy may relate to the large internal variability of the Southern Ocean, but systematic biases are also present in the simulations (Zunz et al. 2013; Turner et al. 2015; Jones et al. 2016). Even the ocean-ice models driven by prescribed forcing derived from atmospheric reanalyses, such as in CORE-II experiments, have trouble reproducing the mean state of the Southern Ocean. For example, CORE-II models display relatively large biases in the position of the ice edge all year long and the CORE-II MMM sea-ice extent is lower than observed, particularly in summer (Farneti et al. 2015; Downes et al. 2015). Part of these common biases are related to the common CORE-II atmospheric forcing.

In addition to sea-ice biases, the majority of the CORE-II models underestimate the MLD in summer while some overestimate it in winter, with a clear impact on the characteristics of the intermediate water masses (Downes et al. 2015). On average, the CORE-II MMM winter mixed layer depth bias is positive and dominated by models with a deep mixed layer and more-saline-than-observed upper ocean. Models with warmer and fresher upper ocean produce shallower-than-observed winter mixed layers. Downes et al. (2015) conclude that the uniformly shallow summer mixed layers are mainly a result of the common atmospheric forcing, while in winter many other additional factors, such as sea ice, surface buoyancy uxes and model parameterisations, a ect the mixed layer depth, and result in varying biases in individual CORE-II models.

Deeper in the ocean, several CORE-II models have cold biases associated with positive MLD biases in the regions of the Antarctic Bottom Water formation. The CORE-II MMM shows warm and saline biases north of 50 S, but cool and fresh biases to the south in the upper 2000 m layer. The fresh bias south of 50 S could be linked to the low levels of brine rejection from ice to the surface ocean related to low CORE-II sea-ice extents (Downes et al. 2015). Below 2000 m depth the CORE-II MMM is biased towards a colder and fresher state than the observational WOA09 climatology.

Inter-ocean exchanges play an important role in global climate in response to variations of local or remote heat and freshwater uxes via the global ocean circulation. This global ocean transport, coupled to global oceanic thermohaline circulation, links the full ocean volume to the climate at long time scales. The Antarctic Circumpolar Current (ACC) is the most intense current of the world ocean and by far the largest conduit for interbasin exchanges.

Farneti et al. (2015) found that the CORE-II MMM Drake passage transport was relatively high (150 Sv), due to two ensemble members, but close to the Climate Model Intercomparison Project Phase 3/5 (CMIP3/5) MMM transport. After excluding these two CORE-II models, the CORE-II MMM transport became closer to observed estimates of 130 to 150 Sv. However, as discussed in Sect. 3.5, CORE-II and CMIP ensembles underestimate more recent ACC transport estimates by Donohue et al. (2016) and de Verdière and Ollitrault (2016).

The CORE-II mass transport time series in the ACC tends to increase during 1948 2008, although this increase attens toward the end of the period. Interestingly, the eddy-permitting models and models with time-dependent and/or three-dimensional eddy-induced coe cients show lower transport trends than the models with a constant or absent

eddy-induced coe cients. This indicates that models which more realistically represent mesoscale eddy e ects do not support long-term increases in the ACC transport, as a response to strengthening westerlies. This ACC insensitivity to the changing winds can be explained by eddy compensation e ects at high resolution and advanced eddy-parameterisation models (Farneti et al. 2015).

These ACC transport trends in CORE-II models are in turn related to the upper ocean water mass structure and linked to temperature, salinity and sea-ice trends. As described by Downes et al. (2015), the CORE-II MMM shows cooling south of 60 S and warming north of the ACC core 50 S in the upper 2000 m. Furthermore, the CORE-II MMM shows a general freshening which, along with the upper ocean temperature trends, can be explained by the stronger and southward moving westerlies which increase the ocean surface heat loss and enhance the atmospheric moisture transport (and therefore the precipitation). Another factor playing a role in the freshening is the redistribution of freshwater by sea ice which is often more important in the Southern Ocean than precipitation (Abernathey et al. 2016; Haumann et al. 2016). These model-produced trends bear good a resemblance to those observed.

For some variables such as the sea-ice concentration, observations with a good spatial coverage are available since 1979 from remote sensing. Despite the uncertainties related to the calibration of the satellite records (e.g. Eisenman et al. 2014), this provides valuable information on the state of the system and an essential metric for model validation. The number of subsurface observations has increased over the last decades thanks to Argo oats (Argo 2000) and sensors attached to marine mammals.

Nevertheless, these observations remain relatively scarce, especially below the sea ice (Schmidtko et al. 2014; de Lavergne et al. 2014; Roemmich et al. 2015; Roquet 2015; Pellichero et al. 2017). The amount of in-situ observations for sea-ice thickness is also relatively limited (Worby et al. 2008). Data assimilation is potentially a powerful tool to obtain estimates for variables that cannot be directly observed or have a poor spatial and temporal observational coverage such as the Antarctic sea-ice thickness (Massonnet et al. 2013), the transport of the subpolar gyres (Duan et al. 2016) and the amount and path of deep water formed close to Antarctica (van Sebille et al. 2013; Azaneu et al. 2014).

3 Material and methods

3.1 Ten selected ocean reanalyses

The ORA output data have been collected in a data base hosted by the Integrated Climate Data Center (ICDC) at Hamburg University[1] and are freely available. Some data were already present from previous ORA-IP studies, but many products were updated and a few new ones added for this study. Ten ORAs were selected to be compared (Table 1), with the most comprehensive temporal overlap over 1993 2010 consisting of all variables required for the diagnostics. The remaining ORAs were discarded due to lack of data either in terms of temporal coverage or variables. Nine ORAs have a global coverage, while one (TOPAZ4) is a regional Arctic-North Atlantic product. Of nine global ORAs, ve are of European origin (all using varying versions of the NEMO ocean), three are American and one is Japanese. All variables analysed were monthly means covering the common intercomparison period from 1993 to 2010 with a few exceptions (mentioned in particular subsections of that diagnostic).

For sea-ice diagnostics, Chevallier et al. (2017) analysed eleven ORAs of which eight are participating in this study, while three (GECCO2, SODA3.3.1 and TOPAZ4) were not previously assessed. Only three ORAs of the other eight (ECDA3, ORAP5 and UR025.4) have not been upgraded meanwhile. As the horizontal resolution of ORAs vary we interpolated all elds onto a common regular 1 latitude-longitude grid for intercomparisons.

Several observational data sets were used to estimate the product-to-observed performance. For the hydrographical analysis, three observational products were used: EN4.2.0.g10 (1993 2010; Good et al. 2013), World Ocean Atlas 2013 (WOA13, 1995 2015; Locarnini et al. 2013; Zweng et al. 2013) and the Sumata Arctic hydrography from Hiroshi Sumata at the Alfred Wegener Institute, Germany based on 1980 2015 observations (Sumata et al. 2017). Notably, the Sumata hydrography is the most comprehensive and up-to-date of the three observational products containing Arctic observations from 28 campaigns from 1980 2015. As for the ORA output, observational data were interpolated onto the common grid for intercomparisons.

3.2 Sea-ice concentration and thickness

Sea-ice concentration (SIC, the relative amount of area covered by ice, compared to some reference area) is the most well-constrained sea-ice variable although not flawless (Ivanova et al. 2014). Satellite observations using passive microwave sensors exist since 1979, available on a daily basis since 1987 at a horizontal resolution ner than 25 km. Chevallier et al. (2017) evaluated various aspects related to sea-ice concentration: the position of the ice edge, sea-ice concentration in the marginal ice zone (concentrations from 15 to 90%) and in the pack ice (concentrations > 90%),

[1] http://icdc.cen.uni-hamburg.de/daten/reanalysis-ocean/oraip.html.

representation of leads within the pack ice, seasonal cycles and trends of integrated Arctic sea-ice area and sea-ice extent.

We use these metrics to evaluate seasonal cycles of sea-ice concentration in both the Arctic and Southern Oceans in the new set of reanalyses. Due to the inclusion of one regional Arctic reanalysis that excludes the North Paci c, the Arctic-integrated extent and area are calculated over a reduced Arctic domain closed at the Bering Strait. We use the same observational datasets as in Chevallier et al. (2017) to assess the realism of ORAs, while taking into account observational uncertainties. Speci cally, these observational sea-ice concentration products are based on the NASATeam algorithm of the National Snow and Ice Data Centre (NSIDC; Cavalieri et al. 1999), from Ifremer/CERSAT using the ARTIST algorithm, and by EUMETSAT Ocean-Sea Ice Satellite Application Facilities (OSISAF). Although these three products have resolutions ner than 25 km, all data are interpolated onto the common regular grid.

Sea-ice thickness (SIT hereafter) is a key diagnostic for assessing the performance of ORAs in the polar oceans. An unrealistic reconstruction of SIT would mean that essential thermodynamic processes controlling ice growth or melt are missing, or that the dynamics of the sea-ice pack is not captured accurately, or both. A major obstacle for the assessment of SIT is the lack of observationally-based data. Unlike sea-ice concentration no large-scale and time-homogeneous records of sea-ice thickness are available.

For the Arctic sea-ice thickness, most of our knowledge relies on collections of datasets from various sources (e.g. Lindsay 2010). Chevallier et al. (2017) used estimates of sea-ice thickness from the ICESat instruments, and estimates of sea-ice volume gathered in Zygmuntowska et al. (2014). In our study, data from the Ice Thickness Regression Procedure (ITRP) are used to analyze the ORA performance. We selected two 2-month periods (February/March and October/November) for the comparison because the ICESat data are available in these months. The ITRP combines upward looking sonar, airborne electromagnetic, NASA operation Icebridge, and ICESat remote sensed ice thickness observations, as explained in detail by Lindsay and Schweiger (2015). Despite the fact that the ITRP thickness data are a result of complex data processing, we believe that the ITRP is the best data set to compare models with. This is due to the following: it allows to calculate sea-ice thickness deviations per grid cell and to integrate total sea-ice volumes in the ITRP region. These metrics are calculated for the period of 2000 2012, with which the ORAs are compared, with the exception of UR025.4 which ends in 2010.

The most comprehensive database adapted for the purpose of evaluating the Antarctic SIT of ORAs is ASPeCt (Worby et al. 2008). This product covers the period 1981 2005 and comprises about 23,000 individual measurements made during ship voyages or helicopter campaigns in the Southern Ocean. Sea-ice thickness was estimated visually by experts onboard. It is therefore likely (1) that systematic errors are present: ships tend to circulate in thin ice, hence estimations are probably biased thin, and (2) that random measurement errors are large, due to the rather simplistic method of measurement (see Worby et al. 2008, for further discussion). The assessment of ORAs with respect to ASPeCt should therefore be conservative and made with extreme caution, in order to not discard ORAs for the wrong reasons.

Unlike the ORA-IP dataset, the ASPeCt data is not gridded and is provided as daily and not monthly values, which complicates further the assessment. We rst binned the ASPeCt data in space and time by matching each of the 23,000 ASPeCt measurements to the corresponding ORA grid cell, year and month over 1993 2005. The number of measurements varies greatly from case to case, but is generally low: in 57% of the cases (one case means one given grid cell during one given month of one given year), less than three measurements are available. We excluded these cases with too few data from our assessment, to limit the probability of detecting a mismatch by chance. For all other cases (four ASPeCt measurements or more in a given month of a given year in a given grid cell), we tested whether the ASPeCt measurements and the ORA-IP monthly mean values could be drawn from the same statistical distribution. For each case, we claimed the ORA product to be compatible with ASPeCt if the ORA estimate fell within the range of all available ASPeCt measurements. In addition we recorded for each case an error equal to the di erence between the reanalysed SIT and the mean value of ASPeCt measurements, and an absolute error equal to the absolute value of the previous metric. The choice of the threshold of at least four ASPeCt measurements to conduct the comparison does not have an impact on the conclusions (not shown here).

Note that Chevallier et al. (2017) carried out a thorough evaluation of the Arctic sea-ice drift in the ORA ensemble, which is not done here for either the Arctic or Antarctic. Sea-ice dynamics is primarily wind driven. Most of the reanalyses considered here use the same atmospheric reanalyses as in the ensemble considered by Chevallier et al. (2017), and there were no signi cant updates in the model physics regarding sea-ice dynamics or rheology. Thus, we can assume that our sea-ice drift results are consistent with those of Chevallier et al. (2017). Hence we refer to their ndings, where necessary.

3.3 Snow depth

Current sea-ice models simulate snow on ice in rather rudimentary ways. Due to its low thermal conductivity and high albedo, snow is strongly altering the snow-ice energy

balance. Both thermal conductivity and albedo depend on the snow density which is kept constant in ORAs (330 to 342 kg m), while observations report a seasonal range of 250 320 kg m from September to May (Warren et al. 1999; Chevallier et al. 2017). Most of the models melt all snow in a grid cell before sea ice is melted at the surface. Many snow related processes (such as precipitation, wind, ice drift and deformation, ooding, melting, evaporation and sublimation) are very uncertain and crudely parameterized in models.

Snow depth observations are very sparse in both polar regions, and in particular in Antarctica. A primary Arctic source is the snow depth climatology of Warren et al. (1999) which is based on data from drifting stations established typically on multi-year sea ice with relatively thick snow cover and collected over the past decades (1954 1991). Due to this, we keep in mind that the Warren climatology is likely overestimating the pan-Arctic average snow depth.

3.4 Mixed layer depth

The oceanic mixed layer constitutes the interface between the atmosphere and the interior of the ocean. This layer is where all dynamic, thermodynamic and biogeochemical air-sea exchanges take place, and where the world s deep water masses acquire their properties (e.g. de Boyer MontØgut et al. 2004; Holte and Talley 2009). As the MLD is a relevant physical index of the vertical mixing intensity in the upper ocean (Toyoda et al. 2017a), the MLDs simulated by the ORAs are evaluated against two observation-based products. These are the Monthly Isopycnal and Mixed-layer Ocean Climatology for the Arctic (MIMOC; Schmidtko et al. 2013) and a recently published Southern Ocean mixed layer climatology (Pellichero et al. 2017).

These products are both based on temperature and salinity pro les from ship observations archived in the World Ocean Database, as well as from oat data from the Argo international program. In addition, MIMOC includes data recorded by ice-tethered pro lers in the Arctic Ocean, while Pellichero et al. (2017) use observations from animal-borne sensor programs in the Southern Ocean (Roquet et al. 2017). These contemporary sources provide an unprecedented data coverage of the sea-ice regions over the entire seasonal cycle. Both climatologies are constructed using an objective mapping of the MLDs computed from instantaneous pro les with the Holte and Talley (2009) algorithm. By contrast, reanalysis MLDs are obtained from monthly mean temperature and salinity elds, using a density threshold of 0.03 kg/m with respect to the value at 10 m depth.

Table 2 Sections used for calculating net lateral volume, heat and freshwater exchange between the Arctic and Sub-Arctic

Section	Latitude		Longitude	
Fram strait	N79		W20	E11
Barents sea opening	N70	N74	E20	
Davis strait	N66		W53	W61
Bering strait	N66		W168	W170

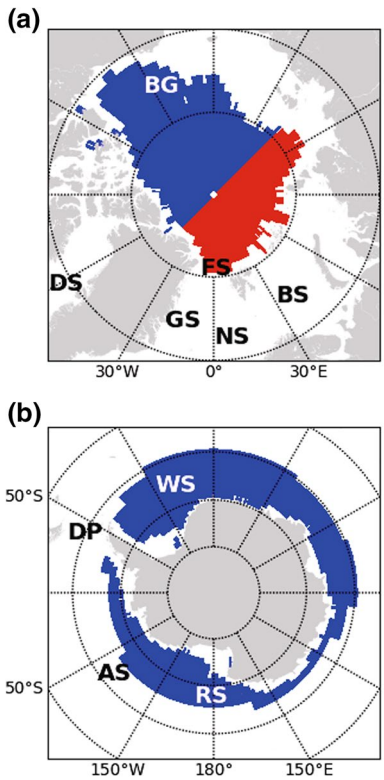


Fig. 1 Regions used to calculate average temperature and salinity pro les. In a red colour shows the region of the Eurasian basin and blue colour the Amerasian basin, while in b blue colour shows the Antarctic open ocean. Annotations: the Fram Strait (FS), Beaufort Gyre (BG), Barents Sea (BS), Davis Strait (DS), Greenland Sea (GS), Norwegian Sea (NS), Amundsen Sea (AS), Ross Sea (RS), Weddell Sea (WS) and Drake Passage (DP)

As noted by de Boyer MontØgut et al. (2004), MLDs computed from monthly, hence smoother, pro les can be underestimated approximately by 10 20 m compared to those based on instantaneous pro les. This is mostly the case in spring when rapid restrati cation occurs (Toyoda et al. 2017a), and needs to be kept in mind when carrying out ORA evaluation. On the other hand, Holte and Talley (2009) found that their algorithm tends to yield slightly shallower MLDs in winter than the density threshold method.

3.5 Liquid ocean transports

Lateral oceanic volume (V), heat (Q), and liquid freshwater transports are calculated through four sections nearly closing the Arctic (see Table 2; Fig. 1). The calculated values represent net transport through the openings, with positive values towards the Arctic. Heat transport is calculated relative to T = 0.1 C (Aagaard and Greisman 1975). Liquid freshwater transport is calculated relative to S = 34.8 on the dimensionless practical salinity scale (Aagaard and Carmack 1989).

Observational ocean transport estimates are obtained from literature, and thus do not represent a consistent time span. Furthermore, their calculations required some assumptions due to discrete spatial sampling of observations. Hence, the observations do not fully close the Arctic Ocean transport budget.

Speci cally, the oceanic ow through the Fram Strait constitutes the main volume and heat exchanges between the Arctic and the Atlantic with a complex re-circulation structure. The total northward ow is estimated as 7 Sv, while a total southward ow of 9 Sv yields a net southward transport of 2 Sv (Table S1; Fahrbach et al. 2001). The heat carried northward along the western coast of Svalbard has shown a relatively large inter-annual variability, between 26 TW (1997/98) and 50 TW (2003/04) (Schauer and Beszczynska-Möller 2009). The ow through the Barents Sea Opening (BSO) towards the Arctic has a net volume ow of 2.3 Sv with about 70 TW heat transport (Table S1; Smedsrud et al. 2013). However, most of this oceanic heat is lost to the atmosphere while en route across the shallow Barents Sea shelf upon reaching the Arctic Ocean (e.g. Gammelsrød et al. 2009).

Another connection between the Arctic and the Atlantic is through the complex channels of the Canadian Arctic Archipelago. However, most of this exchange is channelled through the Davis Strait in Ba n Bay between Greenland and Ba n Island. Here, observations show a net southward volume transport of 1.6 Sv (Table S1; Curry et al. 2014).

The only connection to the Paci c is the shallow Bering Strait. The volume transport through this passage is estimated to be 0.8 Sv directed northward (Table S1; Roach et al. 1995). However, there is a considerable seasonal cycle from 0.4 Sv in winter to about 1.2 Sv in summer (Woodgate and Aagaard 2005), in addition to a possible positive trend in the recent decade (Woodgate et al. 2012). The Bering Strait also represents the only oceanic net freshwater input to the Arctic. Due to its regional Arctic domain, TOPAZ4 model boundary is located in the Bering Strait where a volume transport of 0.7 Sv to the Arctic is prescribed. As temperature and salinity are not prescribed, we decided it is not meaningful to estimate heat and freshwater transports in the Bering Strait for TOPAZ4. Therefore these TOPAZ4 quantities, and consequently the net Arctic heat and freshwater uxes, were excluded from the MMM.

When calculating the ocean transports from the ORA results, the Hudson Strait in the Canadian Arctic Archipelago is omitted, as is the part north of the Barents Sea Opening, i.e., the opening between Bear Island and Spitsbergen Island. These choices make the ORA data more easily compared with the observed transports across the same transects. Some of the modelled ocean transports are calculated based on aggregated data which are interpolated in space and averaged in time, excluding short-term variability. Hence, the ORA data also have some shortcomings with respect to closing budgets for the Arctic Ocean.

For the Southern Ocean transports, we present in Sect. 4.2.3 the values of volume transports across the three main transects of the ACC: the Drake Passage; a transect between South Africa and the Antarctica (Fig. 1, called 30 E); and a transect between Australia and Antarctica (called 147 E). We compare the values estimated from nine global ORAs to estimates from observations.

During the last three decades, the Drake Passage has been more closely monitored than the other two transects. Ganachaud and Wunsch (2000) estimate 140 Sv (– 6 Sv) using an inverse box model applied to WOCE hydrographic data. With a similar method, Lumpkin and Speer (2007) give a mean net transport of 129.7 Sv (– 6.8 Sv). The canonical value of 134 Sv (– 11.2 Sv), obtained by Cunningham et al. (2003) after reviewing ISOS data deployed from January 1979 to February 1980 (Whitworth and R. 1985), is however widely utilized by the physical oceanography community. More recent estimations with a method combining moorings and altimeter 1993 2012 measurements (Koenig et al. 2014) also give a total net transport of 140 Sv (– 10 Sv).

Recent estimations from Donohue et al. (2016), based on 2007 2011 extensive mooring measurements, and from de Verdière and Ollitrault (2016), based on time-mean Argo oat displacements and historical hydrography from the World Ocean Atlas 2009 are likely to be the most reliable ones. Compared to earlier studies, they used methods that reduce uncertainties in the barotropic ow component due to more comprehensive monitoring array and by global mass conserving mean circulation. Donohue et al. (2016) and de Verdière and Ollitrault (2016) provide total transport estimations of 173.3 – 10.7 and 175 Sv, respectively. These values are 30% larger than the canonical value often used as the benchmark for global circulation and climate models.

3.6 Ocean heat and salt contents

Ocean heat and salt contents are denoted as OHC and OSC, respectively. They are calculated as vertical integrals from the reference depth H to the surface :

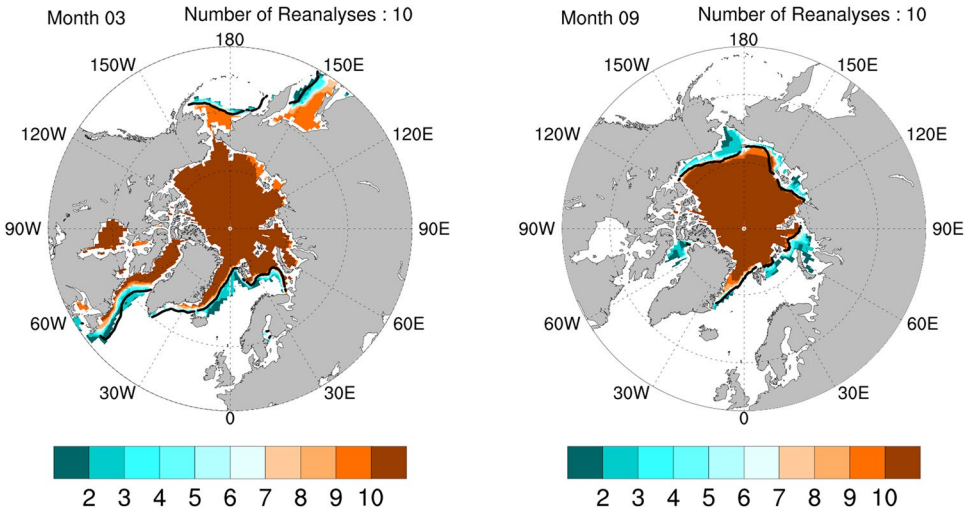


Fig. 2 Number of ORAs per grid cell (up to 10) where their sea-ice concentration is > 15% in March (left) and in September (right) based on 1993 2010 monthly data. Black line is the 15% climatological ice edge by NSIDC NASATeam. The number of reanalyses considered here is 10. Note that the Bering Sea and the Sea of Okhotsk are not a part of the domain of TOPAZ4, so only 9 reanalyses have a solution in these areas

$$v\ i\ d\ \mathrm{m}\qquad \mathrm{d}\boxtimes\boxtimes\boxtimes \tag{1}$$

$$\mathrm{o}\,i \tag{2}$$

where and S are vertical potential temperature and salinity pro les at a horizontal ORA grid point.

The freshwater content, a common oceanographic diagnostic, is the amount of zero-salinity water required to be taken from the ocean or sea ice so that its salinity is changed to the chosen reference salinity and is closely related to OSC and therefore not presented.

3.7 Hydrography

The Antarctic and Arctic ocean basins used to calculate the hydrographic average pro les follow the de nitions given in BarthØlemy et al. (2015). Arctic Ocean was split into two the Eurasian basin and the Amerasian basin, along two meridians, 135 E and 45 W, which join at the North Pole (Fig. 1). The boundary between the two basins approximately follows the Lomonosov Ridge from the East Siberian Shelf to the Lincoln Shelf north of Greenland. The reason for this division of the Arctic Ocean was to see whether product performance varies between the two main Arctic basins, for example in terms of the AW advection.

Due to the vertically integrated ORA-IP hydrographic data only waters located over deep parts of the basins are analysed, analogously to OHC and OSC diagnostics. Speci cally, domain averages are limited by their depth so that in the Arctic the ocean grid points deeper than 500 m are included, while in the Antarctic the limit was 1000 m. The northern limit of the Antarctic basin is chosen as to ensure that the largest fraction of the area is covered with sea ice in winter, and therefore represents a polar marine environment. All ten ORAs and three observational products (Sumata, WOA13 and EN4.2.0.g10) were interpolated to a common 1 horizontal latitude longitude grid, which is identical to the WOA13 grid, before the calculation of regionally averaged hydrographic pro les. As the ORA database does not provide land-sea masks of individual ORAs, we assumed the WOA13 land-sea mask available from the WOA13 website.

First, OHC and OSC for all ORAs were calculated from ve reference depths (H = {100, 300, 700, 1500, 3000 m}) to the surface (= 0 m). After this, the mean potential temperatures and salinities within each layer 100 0 m, 300 100 m, 700 300 m, 1500 700 m and 3000 1500 m were calculated from $a\ \boxtimes\quad L\quad \boxtimes$ as:

$$:1_{,\ \boxtimes} = \ell\ \frac{\boxtimes 1\boxtimes_{,}\quad (\ \boxtimes 1\boxtimes_{\boxtimes}}{,\ (\ \boxtimes} \tag{3}$$

where X is either temperature or salinity, and $G\ _{O}\ \nu$ its average between levels L and U. values where L is deeper than the ocean depth at that particular grid point were excluded from the further analysis. Finally, level averaged temperatures and salinities $1_{,\ :2}$ were temporally and basin-averaged.

4 Results

4.1 Arctic mean states

4.1.1 Sea ice and snow

Ten ORAs show an overall agreement in the location of the sea-ice edge in the Arctic Ocean and along its margins (Figs. 2, S1 and S2), which can be attributed to sea-ice data assimilation and the constraint by the atmospheric forcing. On average, there is a good agreement with respect to the

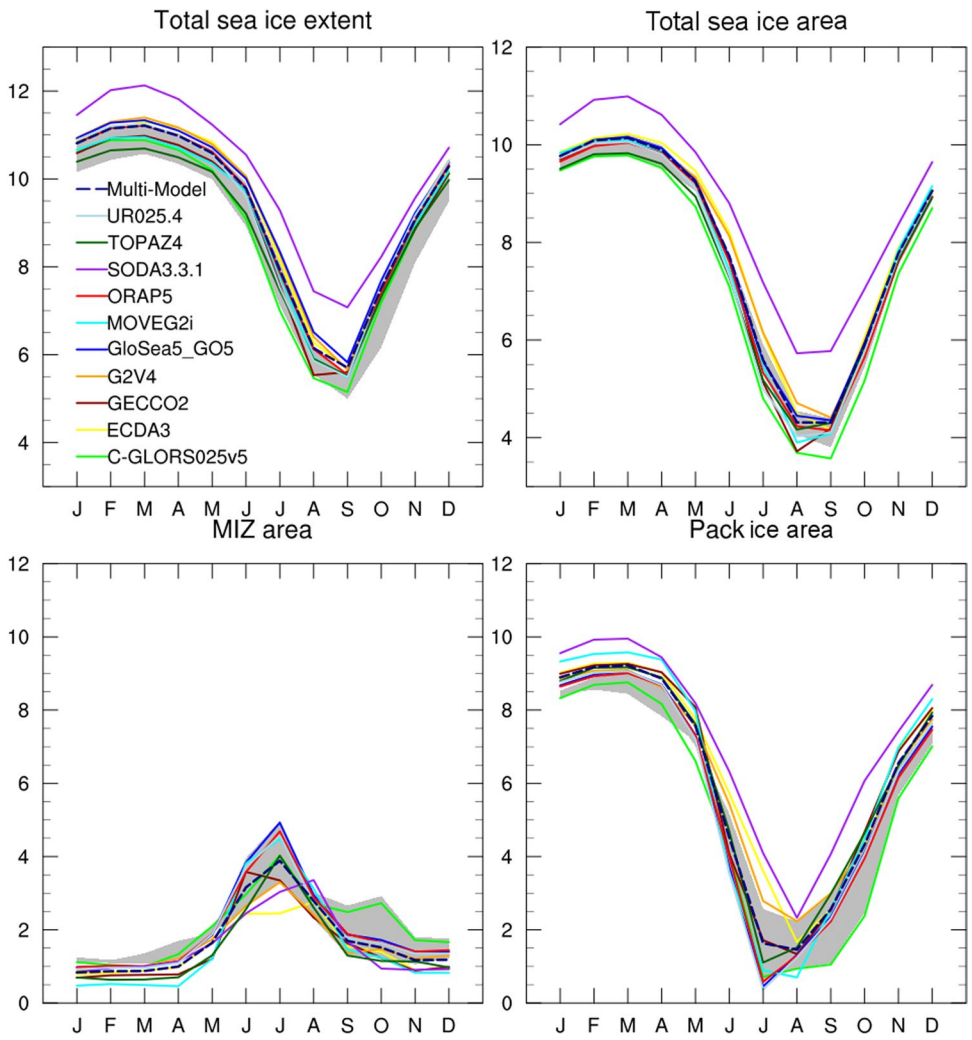


Fig. 3 Mean seasonal cycle (over the period of 1993 2010) of the Arctic sea-ice extent and area (upper row), and of the area covered by Marginal Ice Zone (MIZ) and pack ice (lower row), in all ORAs (colour lines) and in NSIDC, CERSAT and OSISAF observations (grey shading). Domain of integration excludes the ocean area in the North Paci c south of Bering Strait. MIZ is de ned as a region where the sea-ice concentration is less than 90% and greater than 15%, while the pack ice is the region where the sea-ice concentration is higher than 90%. Units are in 10^C km

sea-ice edge in the Barents Sea, the Greenland Sea and the Bering Sea. Most reanalyses lack sea ice in the Labrador Sea and the Sea of Okhotsk, as Chevallier et al. (2017) pointed out. A few ORAs simulate too much sea ice eastward of the coasts of the Labrador Sea and the Greenland Sea: these are the ORAs that do not assimilate sea-ice concentration (Table 1; Figure S1). In summer, a number of ORAs underestimate the presence of sea ice east of Greenland, and some underestimate sea-ice melt near the shelves, in the Kara Sea and in Ba n Bay.

Figure 3 shows the seasonal cycles of Arctic sea-ice extent and area in ten ORAs. The modeled seasonal cycle is generally in phase with observations, with a maximum (minimum) sea-ice area and extent in March (September), although a few ORAs simulate sea-ice extent minima in August. SODA3.3.1 overestimates sea-ice extent and area in all months, so it is excluded from the subsequent Arctic sea-ice concentration ensemble analysis. The ensemble spread of ORA sea-ice extent, without SODA3.3.1, is limited over the year, and is comparable to the estimated observational uncertainty. This was expected, since most reanalyses assimilate sea-ice concentration. The spread is larger during the winter months, and all ORAs align well during refreezing in autumn. A few ORAs exhibit systematic biases compared to the observations in the winter months, which is consistent with the lack of sea ice in the Labrador Sea, as noted above. In most ORAs, the simulated August September sea-ice extents are within the observational uncertainty. Results are similar for sea-ice area, although its ensemble spread is larger in spring and summer than the sea-ice extent spread. For both sea-ice extent and area, the MMM mean without SODA3.3.1 is near the upper range of the observational estimates.

The signi cant spread in sea-ice area denotes di erences in the distribution of sea-ice concentration within the ice cover. As in Chevallier et al. (2017), we investigate the separate contributions of Marginal Ice Zone (MIZ) and pack ice in the total area spread. In the observations, the MIZ

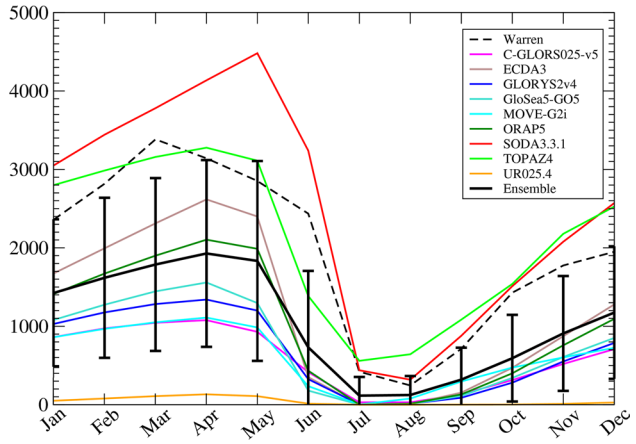


Fig. 4 Monthly climatology of the Arctic snow volume (km) of the ORA-IP models for the period from 1993 to 2010, its ensemble mean (black solid line errors bars designate one standard deviation uncertainty) and the snow volume of the Warren climatology (black dashed line). The snow volume is calculated for the entire Arctic Ocean excluding regions south of the Fram Strait

area varies between 1 and 2 million km^{z} from November to April, peaks in July, and decreases slowly from August to October (Fig. 3). Three observational products give consistent results, although CERSAT has a systematically smaller MIZ area in June September. During October December, the spread among the observational estimates is the largest, when NSIDC has a larger MIZ than the others. The pack-ice area has a seasonal cycle evolving at the same rate as total sea-ice area, although its annual minimum is reached in July August. In the Arctic Ocean, sea ice is predominantly pack ice, except in summer when the MIZ/pack-ice area ratio is over 50%.

The ORAs reproduce these seasonal sea-ice extent and area cycles relatively well. Most ORAs are consistent with the ice product they assimilate (e.g. C-GLORS025v5 with NSIDC, GLORYS2v4 with CERSAT; Table 1). However, during winter and early spring, all ORAs simulate MIZ area lower than observed, and systematically too high pack-ice area when the assimilated ice product is taken into account (lower right panel of Fig. 3). In summer, the ensemble spread is larger, and there are a number of ORAs that align well with observational estimates. But no ORA simulates more MIZ than observed, and a few ORAs stand out with a lower-than-observed MIZ peak area: those are the products without data assimilation (Table 1). They tend to simulate very high sea-ice concentration almost all year long (not shown).

The snow volume in the ORAs varies widely not only between the ORAs using di erent precipitation data sets but also between the ORAs using ERA-Interim precipitation rates (Fig. 4; Table 1). As apparent from Figs. 4 and S3, ORAs have a thinner snow cover everywhere in the Arctic and hence smaller snow volumes than Warren et al. (1999), which is known to have a thick bias, as explained earlier (Figs. 4, S3). The maximum snow volume in the Warren climatology occurs between March and April with values around 3000 km . The ORA values range between > 4000 km (SODA3.3.1) and < 200 km$^{:}$ (UR025.4). By inspecting the ORA ensemble mean and its standard deviation we can identify three ORAs which deviate most from the other ORAs: UR025.4 which has almost no snow at all, SODA3.3.1, driven by the MERRA2 reanalysis and associated with a high bias in sea-ice area, which exceeds the Warren climatology for all months, and TOPAZ4 which ts very closely to the Warren climatology, despite being driven by ERA-Interim. The remaining ORA snow volumes range from about 1000 km (MOVE-G2i) to 2500 km^{t} (ECDA3). The large variation between the ORAs driven by the same reanalysis (ERA-Interim) is surprising. This might point to large uncertainties in process parameterisations (related to for example sea-ice ridging and sublimation) which alter the snow depth.

All ORAs show a strong decrease of the snow volume from May to June (Fig. 4). This is certainly connected to the fact that ORAs rst have to melt all snow o before their sea ice starts to melt. Related to this, all ORAs except SODA3.3.1 and TOPAZ4 have almost no snow on ice from July to August. Then from September to December the majority of ORAs (except UR025.4, SODA3.3.1 and TOPAZ4) show only moderate di erences in the snow volume. Interestingly, di erences between the ORA snow volumes grow strongly from January to April.

The mean di erence of the sea-ice thickness of the ORAs relative to the ITRP data for February March is presented in Fig. 5. Most ORAs underestimate the ice thickness north of the Canadian Arctic Archipelago, north of Greenland and the Fram Strait. Especially large deviations are found for ECDA3, MOVE-G2i, SODA3.3.1, and UR025.4 for which the deviations can amount to more then 2 m. More moderate deviations are detected for C-GLORS025v5, GECCO2, GloSea5-GO5, and TOPAZ4. ORAP5 exhibits only a minor underestimation while GLORYS2v4 overestimates the ice thickness by up to 1 m. In the Beaufort Sea, some of the ORAs overestimate the ice thickness moderately (C-GLORS025v5, GloSea5-GO5, SODA3.3.1) while ORAP5 exceeds the observed thickness by up to 1 m and GLORYS2v4 by up to 2 m. TOPAZ4 and GECCO2 show no notable deviations in the Beaufort Sea. Most of the ORAs overestimate the thickness over the Eurasian shelves. GLORYS2v4 strongly overestimates ice thickness over almost the whole Arctic Ocean. In October November, the ORA-ITRP mean di erences generally appear similar to the di erences in February March, but with a tendency towards larger underestimations of sea-ice thickness (Figure S4).

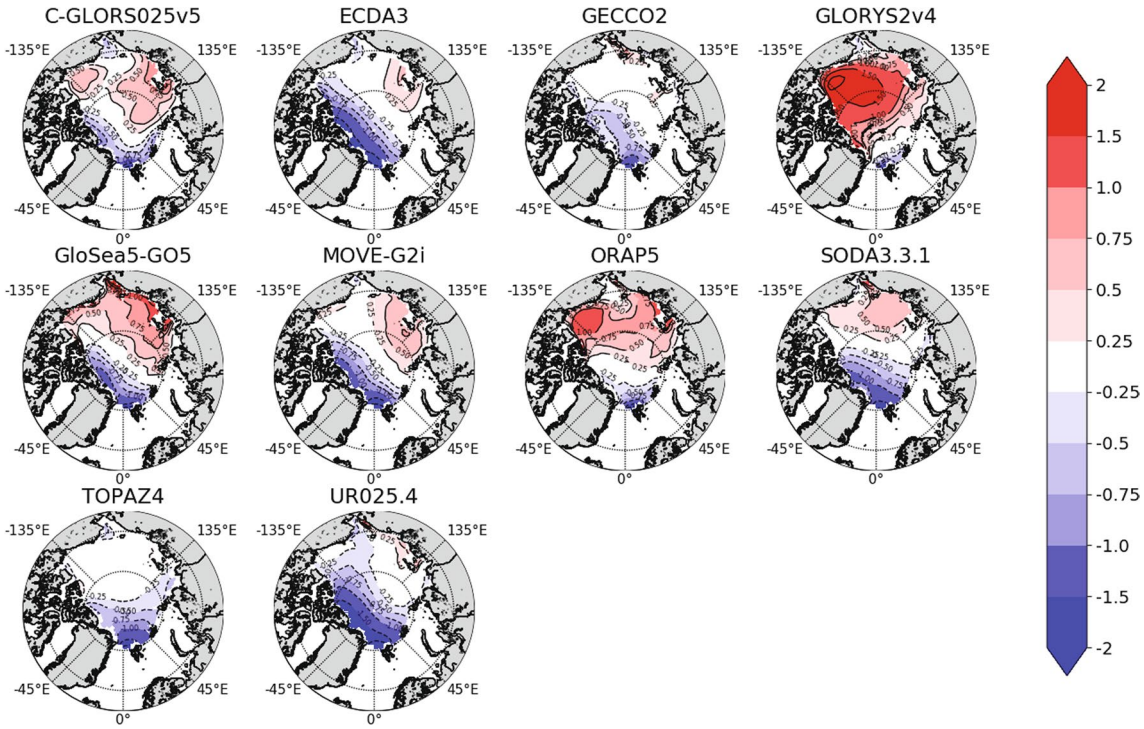


Fig. 5 The 2000 2012 mean di erence of the ORAs to the ITRP sea-ice thickness (m) in February March

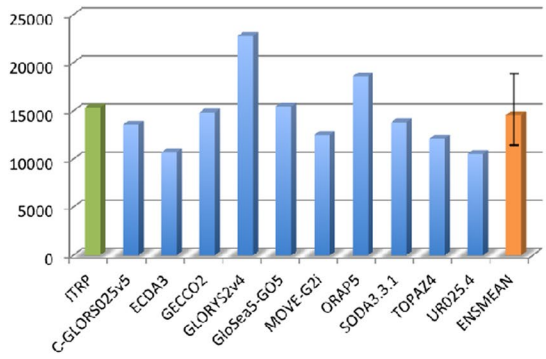
In February March the mean (period 2000 to 2012) ice volume amounts in ITRP to 15,400 km (Fig. 6a). Corresponding ORA ice volumes range between 10,500 and 12,800 km$^{\backprime}$ with the ensemble mean of 14,500 km . Two ORAs are very close to the ITRP value (GECCO2 and GloSea5-GO5), but this is, at least in the case of GloSea5-GO5, due to compensating regional biases (Fig. 5). In October November the mean ITRP ice volume is about 12,400 km^{d}, while the ORA range is large, from 5300 to 19,200 km (Fig. 6b). The average ice volume of ve ORAs (C-GLORS025v5, ECDA3, GloSea5-GO5, MOVE-G2i and UR025.4) stays low below 8000 km . Correspondingly, the ORA MMM ice volume is much lower than the ITRP value (about 10,000 km).

Figure 6c displays the mean sea-ice volume loss between February March and October November calculated in the ICESat domain (i.e. the di erence between Fig. 6a, b). While the ITRP seasonal volume loss is about 3000 km^{C}, seasonal volume losses in six ORAs exceed or are close to 5000 km , indicating too high seasonal sea-ice volume amplitudes. Accordingly the ORA MMM volume loss is biased high (4500 km^{C}).

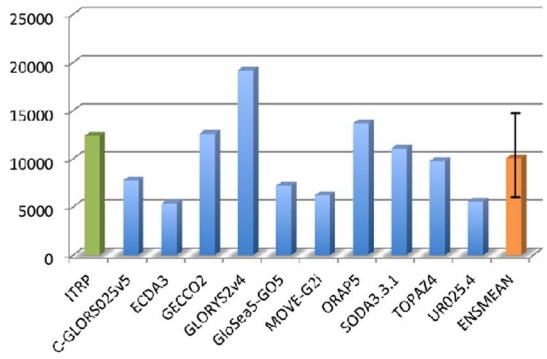
4.1.2 Mixed layer depth

There is no systematic bias in the representation of winter MLDs in the Arctic Ocean in the various ORAs (Figs. 7, S5). UR025.4, ECDA3, GloSea5-GO5 and GLORYS2v4 give the largest MLD overestimations, while MOVE-G2i and GECCO2 yield the strongest underestimations. The observed pattern, with MLDs around 40 m in the Amundsen and Makarov Basin and with shallower mixed layers in the Amerasian Basin, is not closely matched by any of the products. In the Barents Sea and south of Svalbard, all ORAs simulate deeper mixed layers than in the observation-based product (Figure S5). The di erence between the ensemble mean and the climatology exceeds 400 m locally.

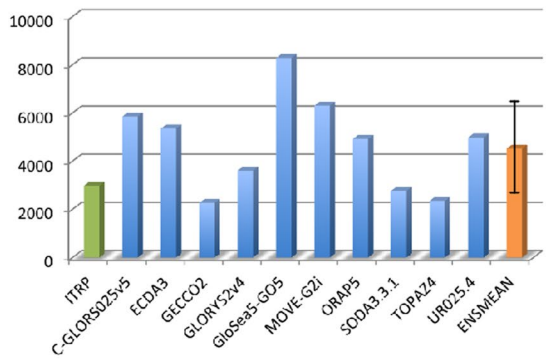
In summer, all ORAs underestimate MLDs (Figure S6). These shallow mixed layers are generally not due to the coarse vertical grid, because the top-level thicknesses of ocean models are mostly from 1 3 m, only ECDA3, GECCO2 and SODA3.3.1 have the top-layer thickesses of 10 m (Table 1). The ensemble mean bias reaches as much as 20 m under sea ice (Fig. 7), although the reliability of the climatology might be questioned in the regions just north of the Canadian Arctic Archipelago and Greenland, where the ice is very thick and few hydrographic measurements exist. As a result, the MMM negative bias is of the order of 10 m in the Barents Sea and is smaller in the Greenland and Norwegian Seas.



(a) Mean Feb/Mar



(b) Mean Oct/Nov



(c) Loss Feb/Mar to Oct/Nov

Fig. 6 The 2000 2012 mean sea-ice volume (km) in the ICESat domain in a February March and b October November. c Mean ice volume loss km^0 in the ICESat domain between February March and October November (the di erence of a and b). The ORAs are denoted by blue bars, the ITRP by green bars and the ensemble mean (ENSMEAN) by orange bars. The error bar in ENSMEAN represents the ORA ensemble spread (standard deviation)

4.1.3 Liquid ocean transports

The net oceanic exchange between the Arctic Ocean and the sub-Arctic seas through the four major openings, the Fram Strait, Bering Strait, Davis Strait, and BSO, from ORAs and observations are summarized in Table S1 and illustrated as bar plots in Figs. 8, S7 and S8. Generally, ORA volume transports are within the observational uncertainty and for the BSO also heat transports, yet with some notable di er-ences. All ORAs close the Arctic Ocean volume transport budget comparably to the observations. All ORAs tend to be on the low side in terms of net oceanic heat transport towards the Arctic (Fig. 8). This could imply a negative temperature bias in the ORAs. Indeed, in the Barents Sea, WOA13 and EN4.2.0.g10 are warmer than the ORA MMM, but Sumata is slightly cooler than the MMM (Figure S9). GECCO2 is an exception which also shows excessive heat transports through the BSO. Most of the ORAs also tend to be on the low side with respect to heat transport through the Fram Strait due to either too cold northward owing AW or southward recirculation of too warm AW. However, some ORAs are within the uncertainty range of the observations. A caveat to this analysis is our method for computation of the heat transports. For comparison with observation-based estimates from literature, we have chosen to compute the heat transports based on net volume transports through each separate section and relative to a xed reference tempera-ture (T { } C; Aagaard and Greisman 1975), which is also the common method used in the literature. However, this method has some inherent inconsistencies related to the lack of a closed volume transport budget and the actual temperature di erence between the incoming and outgoing water masses. A more consistent method for the computa-tion of ocean heat transports is discussed in Schauer and Beszczynska-Möller (2009).

The ORAs show a generally good agreement with observations with respect to freshwater export from the Arctic, except GLORYS2v4 and MOVE-G2i, which have

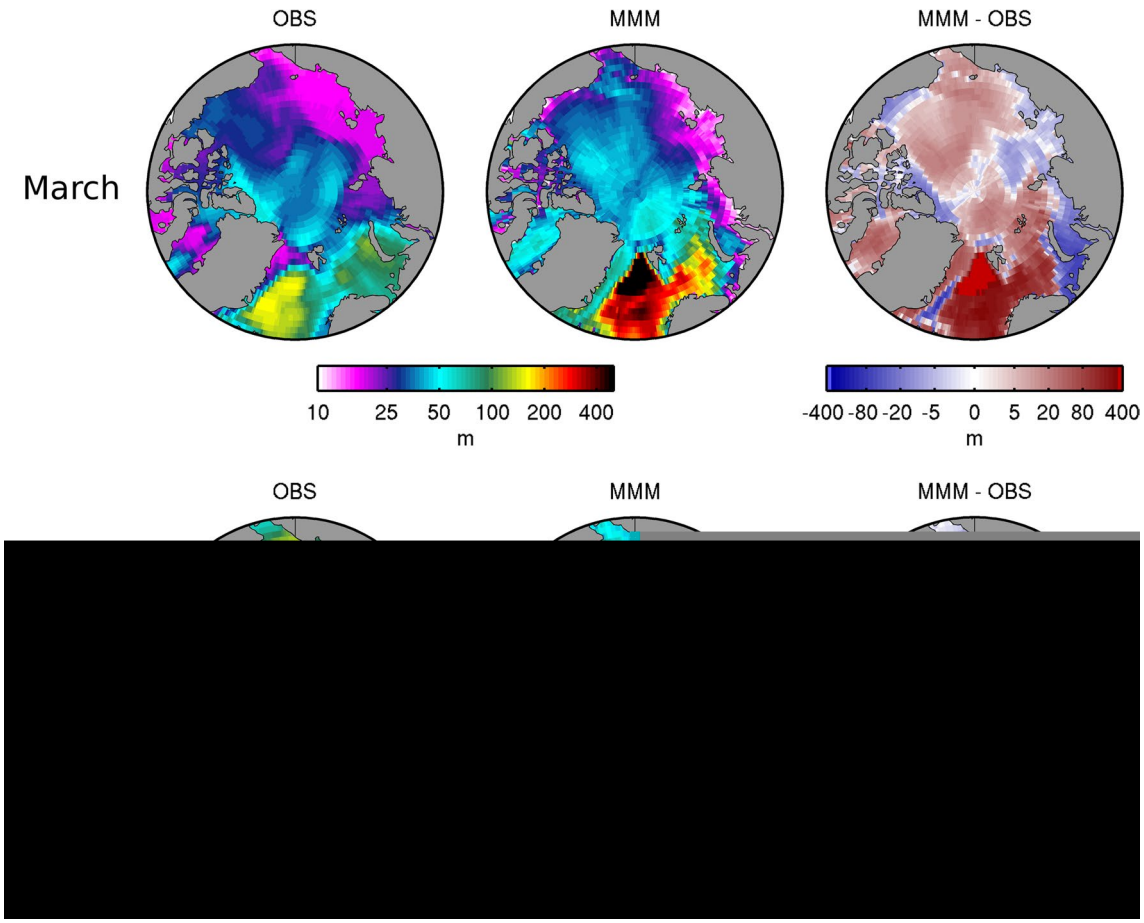


Fig. 7 Arctic mixed layer depth for March (top) and August (bottom), in the MIMOC climatology (OBS; Schmidtko et al. 2013), for the ORA ensemble mean (MMM) and the bias of the ensemble mean with respect to the climatology (MMM OBS). Note the logarithmic colour scale used for March elds

particularly low freshwater exports (Table S1; Figure S8). The reason for this discrepancy is not clear, but GLORYS2v4 has a positive salinity bias in the Arctic Ocean (see Sect. 4.1.4). On the other hand, C-GLORS025v5 shows a good agreement with observations in terms of total Arctic freshwater budget, but it has a di erent distribution with low freshwater volumes exported through the Fram Strait and an enhanced export through the Davis Strait.

The MMM represents an estimate of Arctic sub-Arctic exchanges comparable to observed estimates (Table S1; Figure S8). The MMM freshwater transport through the Bering and Davis straits are in close agreement with the observations, while the MMM freshwater transport through the Fram Strait is on the low side, although the volume transport is comparable to the observations. Overall, the MMM is generally in closer agreement with the observations than individual ORA estimates.

In terms of heat transport variability, represented by one standard deviation based on annual averages, all ORAs overlap with the observed range of variability in the main

Fig. 8 Mean 1993 2010 Arctic liquid ocean heat transport through the major Arctic Ocean openings, and their net sum (total) in TW. Error bars represent standard deviations of monthly values. See Supplementary Table S1 for values of the error bars and references to the calculation of error bars in the observations